EXHIBIT B: CLEAN VERSION OF THE PENDING CLAIMS
U.S. APPLICATION SERIAL NO. 09/220,142
(ATTORNEY DOCKET NO. 9301-035-999)

(as amended July 30, 2001)

1. (Four Times Amended) A method of determining a consensus profile for a first plurality of drug perturbations to a cell type or organism, said method comprising identifying among a plurality of sets of cellular constituents in a plurality of response profiles one or more sets of cellular constituents, each of said one or more sets of cellular constituents being upregulated or downregulated by said first plurality of drug perturbations, each response profile in said plurality of response profiles (i) comprising measurements of a plurality of cellular constituents, and (ii) resulting from a different drug perturbation to said type of cell or organism, wherein each set of cellular constituents in said plurality of sets of cellular constituents consists of cellular constituents that co-vary under a second plurality of perturbations or that are co-regulated, wherein said plurality of response profiles comprises at least five response profiles, and wherein said consensus profile for said first plurality of drug perturbations comprises measurements of said one or more sets of cellular constituents.

3. (Amended) The method of claim 1, wherein the plurality of response profiles comprises more than ten response profiles.

4. The method of claim 3, wherein the plurality of response profiles comprises more than 50 response profiles.

5. The method of claim 4, wherein the plurality of response profiles comprises more than 100 response profiles.

6. (Three Times Amended) The method of claim 1, wherein said first plurality of drug perturbations is associated with a particular biological effect.

7. The method of claim 6, wherein the particular biological effect is the effect of a particular class or type of drug.

8. The method of claim 6, wherein the particular biological effect is a therapeutic effect.

9. The method of claim 6, wherein the particular biological effect is a toxic effect.

10. (Twice Amended) The method of claim 1, wherein each of the sets of cellular constituents consists of cellular constituents which are co-regulated.

11. (Twice Amended) The method of claim 1, wherein each of the sets of cellular constituents consists of cellular constituents which co-vary in the plurality of response profiles.

12. The method of claim 11, wherein the cellular constituents which co-vary are identified by cluster analysis of cellular constituents in the plurality of response profiles.

13. The method of claim 12, wherein the cluster analysis is done by means of a clustering algorithm.

14. The method of claim 13, wherein the clustering algorithm is *hclust*.

15. The method of claim 12, wherein said cluster analysis determines a clustering tree, the cellular constituents which co-vary comprising branches of said clustering tree.

16. The method of claim 15, wherein the sets of co-varying cellular constituents are selected from a branching level of the clustering tree.

17. The method of claim 12, wherein a statistical significance for the sets of co-varying cellular constituents is determined by means of an objective statistical test.

18. The method of claim 17, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in cluster analysis of the cellular constituents;

(b) generating permuted response of cellular constituents by means of Monte Carlo randomization of perturbation index for the response of each cellular constituent across all perturbations;

(c) performing cluster analysis on the permuted response of cellular constituents;

(d) determining the fractional improvement in the cluster analysis on the permuted response of cellular constituents; and

(e) repeating said steps of generating permuted response of cellular constituents and performing cluster analysis on the permuted response of cellular constituents so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

19. (Amended) The method of claim 1, wherein the one or more sets of cellular constituents are identified by re-ordering the response profiles into sets associated with similar biological effects.

20. The method of claim 19, wherein the sets of response profiles associated with similar biological effects are identified by cluster analysis of the response profiles.

21. The method of claim 20, wherein the cluster analysis is done by means of a clustering algorithm.

22. The method of claim 21, wherein the clustering algorithm is *hclust*.

23. The method of claim 20, wherein said cluster analysis determines a clustering tree, the response profiles associated with similar biological effects comprising branches of said clustering tree.

24. The method of claim 23, wherein the branches are selected by applying a cutting level across said clustering tree, said cutting level being determined by an expected number of biological pathways represented by the sets of cellular constituents.

25. The method of claim 20, wherein a statistical significance for the sets of response profiles is determined by means of an objective statistical test.

26. The method of claim 25, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in the cluster analysis of the response profiles;

(b) generating permuted response profiles by means of Monte Carlo randomization of cellular constituent index for each response profile across the measured cellular constituents;

(c) performing cluster analysis on the permuted response profiles;

(d) determining the fractional improvement in the cluster analysis on the permuted response profiles; and

(e) repeating said steps of generating permuted response profiles and performing cluster analysis on the permuted response profiles so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

27. The method of claim 1, wherein the sets of cellular constituents are basis cellular constituent sets.

28. The method of claim 27, wherein the basis cellular constituent sets are genesets.

29. (Twice Amended) A method of determining a consensus profile for a first plurality of perturbations to a cell type or organism, said method comprising identifying among a plurality of sets of cellular constituents in a plurality of projected profiles one or more sets of cellular constituents, each of said one or more sets of cellular constituents being upregulated or downregulated by said first plurality of perturbations, each projected profile in said plurality of projected profiles

(i) resulting from a different perturbation to said type of cell or organism, and

(ii) comprising measurements of a plurality of cellular constituents in said type of cell or organism that have been projected onto basis cellular constituent sets, said basis cellular

constituent sets being defined by co-variation of measurements of cellular constituents under a second plurality of different perturbations, wherein said consensus profile for said first plurality of perturbations comprises projected measurements of said one or more sets of cellular constituents.

30. (Three Times Amended) The method of claim 1 wherein the consensus profile is the intersection of the sets of cellular constituents activated or de-activated by said first plurality of drug perturbations.

31. (Twice Amended) The method of claim 29, wherein the consensus profile is the intersection of the sets of cellular constituents activated or de-activated by said first plurality of perturbations.

32. (Amended) The method of claim 30 or 31, wherein the one or more sets of cellular constituents are identified by re-ordering the response profiles into sets associated with similar biological effects.

33. The method of claim 31, wherein the intersection is identified by visual inspection of the plurality of projected response profiles.

34. The method of claim 32, wherein the intersection is identified by visual inspection of the plurality of projected response profiles.

35. The method of claim 31, wherein the intersection is identified by thresholding the projected response profiles.

36. The method of claim 31, wherein the intersection is identified arithmetically.

37. The method of claim 36, wherein the intersection is identified by a method comprising:

(a) replacing amplitudes of cellular constituent sets in the projected response profiles that are above a threshold with values of unity;

(b) replacing amplitudes of cellular constituent sets in the projected response profiles that are below said threshold with values of zero; and

(c) determining the element-wise product of the projected response profiles, wherein the element-wise product of the projected response profiles is the intersection.

38. (Three times Amended) A method of determining a consensus profile for a first plurality of perturbations to a cell type or organism, said method comprising identifying among a plurality of sets of genes in a plurality of response profiles one or more sets of genes, each of said one or more sets of genes being upregulated or downregulated by said first plurality of perturbations, each response profile in said plurality of response profiles (i) comprising measurements of transcript levels for a plurality of genes, and (ii) resulting from a different perturbation to said type of cell or organism, wherein each set of genes in said plurality of sets of genes consists of genes having transcripts that co-vary under a second plurality of perturbations or that are co-regulated, and wherein said consensus profile for said perturbations comprises measurements of transcript levels for said one or more sets of genes.

39. (Twice Amended) A method for comparing a biological response profile to a consensus profile, said consensus profile comprising projected measurements of one or more sets of cellular constituents, said one or more sets having been identified among a plurality of sets of cellular consituents in a plurality of projected response profiles, each of said one or more sets of cellular constituents being upregulated or downregulated by a first plurality of perturbations, each projected response profile in said plurality of projected response profiles

(i) resulting from a different perturbation to said type of cell or organism, and

(ii) comprising measurements of a plurality of cellular constituents in said type of cell or organism that have been projected onto basis cellular constituent sets, said basis cellular constituent sets being defined by co-variation of measurements of cellular constituents under a second plurality of different perturbations, said method comprising:

(a) converting the biological response profile into a projected response profile by projecting measurements of cellular constituents in said biological response profile onto said basis cellular constituent sets; and

(b) determining the value of a similarity metric between the projected response profile and the consensus profile.

40. The method of claim 39, wherein said step of converting comprising projecting the biological response profile onto the basis cellular constituent sets.

41. The method of claim 39, wherein the similarity metric is the generalized cosine angle between the projected response profile and the consensus profile.

42. The method of claim 39, further comprising a step of determining the statistical significance of the similarity metric.

43. The method of claim 42, wherein the statistical significance is assessed using an empirical probability of distribution generated under a null hypothesis of no correlation.

44. (Four Times Amended) A method for grouping measured response profiles in sets which are associated with similar biological effects comprising grouping response profiles among a plurality of response profiles into sets, each of said sets of response profiles consisting of response profiles in which the responses of one or more sets of genes in each response profile are similar among response profiles in the set, each response profile in said plurality of response profiles (i) comprising measurements of transcript levels of a plurality of genes, and (ii) resulting from a different perturbation, wherein each of said sets of genes consists of genes that co-vary under a plurality of perturbations or that are co-regulated, wherein said plurality of response profiles comprises at least five response profiles.

45. The method of claim 44, wherein the sets of response profiles are identified by cluster analysis of the response profiles.

46. The method of claim 45, wherein the cluster analysis is done by means of a clustering algorithm.

47. (Amended) The method of claim 46, wherein the clustering algorithm is *hclust*.

48. The method of claim 45, wherein said cluster analysis determines a clustering tree, the sets of response profiles comprising branches of said clustering tree.

49. The method of claim 45, wherein a statistical significance for the sets of response profiles is determined by means of an objective statistical test.

50. (Amended) The method of claim 49, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in the cluster analysis of the response profiles;

(b) generating permuted response profiles by means of Monte Carlo randomization of gene index for each response profile across the measured genes;

(c) performing cluster analysis on the permuted response profiles;

(d) determining the fractional improvement in the cluster analysis of the permuted response profiles; and

(e) repeating said steps of generating permuted response profiles and performing cluster analysis on the permuted response profiles so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

58. (Twice Amended) A method for determining the therapeutic efficacy of a drug or drug candidate comprising identifying one or more groups of sets of cellular constituents in one or more response profiles associated with exposure to the drug or drug candidate, each response profile comprising measurements of a plurality of cellular constituents, wherein each of said groups is indicative of a particular therapeutic effect, and wherein the therapeutic effect of the drug or drug candidate is determined to be the particular therapeutic effect indicated by the identified groups, wherein each of said sets of cellular constituents consists of cellular constituents that co-vary under a plurality of perturbations or that are co-regulated.

59. The method of claim 58, wherein the sets of cellular constituents are determined by a method comprising performing cluster analysis of the response profiles.

60. The method of claim 59, wherein the cluster analysis is done by means of a clustering algorithm.

61. The method of claim 60, wherein the clustering algorithm is *hclust*.

62. The method of claim 59, wherein said cluster analysis determines a clustering tree, the sets of cellular constituents comprising branches of said clustering tree.

63. The method of claim 59, wherein a statistical significance for the sets of cellular constituents is determined by means of an objective statistical test.

64. The method of claim 63, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in the cluster analysis of cellular constituents;

(b) generating permuted response of cellular constituents by means of Monte Carlo randomization of the perturbation index for each cellular constituent across all perturbations;

(c) performing cluster analysis on the permuted response of cellular constituents;

(d) determining the fractional improvement in the cluster analysis of the permuted response of cellular constituents; and

(e) repeating said steps of generating permuted response of cellular constituents and performing cluster analysis on the permuted response of cellular constituents so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

72. (Three Times Amended) A method for analyzing response data from a biological sample comprising

(a) grouping cellular constituents from the biological sample into sets of genes that co-vary in a plurality of response profiles, each response profile in said plurality of response profiles (i) comprising measurements of transcript levels

of a plurality of genes, and (ii) resulting from a different perturbation to said biological sample; and

(b) grouping the plurality of response profiles into sets of response profiles that similarly affect genes,

wherein said plurality of response profiles comprises at least five response profiles.

73. (Amended) The method of claim 72, wherein one or more genes which co-vary in association with a particular biological effect are identified from the sets of genes that co-vary in said plurality of response profiles.

74. (Amended) The method of claim 72, wherein one or more response profiles that are associated with a particular biological effect are identified from the sets of response profiles that similarly affect genes.

75. The method of claim 73 or 74, wherein the particular biological effect is an effect on a biological pathway.

76. (Amended) The method of claim 73, wherein one or more genes associated with said biological effect are identified.

77. The method of claim 76 wherein the one or more genes identified comprise known genes.

78. The method of claim 76, wherein the one or more genes identified comprise previously unknown genes.

89. The method of claim 1, wherein said sets of cellular constituents are co-varying cellular constituent sets.

90. The method of claim 89, wherein the cellular constituents which co-vary are identified by cluster analysis.

91. The method of claim 89, wherein the cluster analysis is done by means of a clustering algorithm.

92. The method of claim 91, wherein the clustering algorithm is *hclust*.

93. The method of claim 90, wherein said cluster analysis determines a clustering tree, the cellular constituents which co-vary comprising branches of said clustering tree.

94. The method of claim 93, wherein the sets of co-varying cellular constituents are selected from a branching level of the clustering tree.

95. The method of claim 90, wherein a statistical significance for the sets of co-varying cellular constituents is determined by means of an objective statistical test.

96. The method of claim 95, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in cluster analysis of the cellular constituents;

(b) generating permuted response of cellular constituents by means of Monte Carlo randomization of the perturbation index for response of each cellular constituent across the set of perturbations;

(c) performing cluster analysis on the permuted response of cellular constituents;

(d) determining the fractional improvement in the cluster analysis on the permuted response of cellular constituents; and

(e) repeating said steps of generating permuted response of cellular constituents and performing cluster analysis on the permuted response of cellular constituents so that a distribution of fractional improvements is obtained,

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

97. The method of claim 39, 40, 41, 42, or 43, wherein said sets of co-varying cellular constituents comprise cellular constituents which co-vary in the plurality of response profiles.

98. The method of claim 72, wherein step (a) is carried out before step (b).

99. The method of claim 72, wherein step (b) is carried out before step (a).

100. (Three Times Amended) A method of grouping sets of drug perturbations that similarly affect cellular constituents in a cell type or organism among a plurality of drug perturbations comprising grouping response profiles among a plurality of response profiles in sets, each of said sets of response profiles consisting of response profiles in which the responses of one or more sets of cellular constituents are similar among the response profiles in the set, each response profile in said plurality of response profiles (i) comprising measurements of a plurality of cellular constituents, and (ii) resulting from a different drug perturbation, wherein each of said sets of cellular constituents consists of cellular constituents that co-vary under a plurality of perturbations or that are co-regulated, thereby grouping said sets of drug perturbations, wherein said plurality of response profiles comprises at least five response profiles.

105. (Amended) A method for grouping measured response profiles in sets which are associated with similar biological effects comprising grouping response profiles in sets among a plurality of response profiles by cluster analysis of said plurality of response profiles, said sets of response profiles consisting of response profiles having similar responses of a group of cellular constituents, each response profile in said plurality of response profiles (i) comprising measurements of a plurality of cellular constituents, and (ii) resulting from a different perturbation, wherein a statistical significance for the sets of response profiles is determined by means of an objective statistical test.

106. The method of claim 105, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in the cluster analysis of the response profiles;

(b) generating permuted response profiles by means of Monte Carlo randomization of cellular constituent index for each response profile across the measured cellular constituents;

(c) performing cluster analysis on the permuted response profiles;

(d) determining the fractional improvement in the cluster analysis of the permuted response profiles; and

(e) repeating said steps of generating permuted response profiles and performing cluster analysis on the permuted response profiles so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

107. (New) The method of claim 38, wherein the plurality of response profiles comprises more than five response profiles.

108. (New) The method of claim 107, wherein the plurality of response profiles comprises more than ten response profiles.

109. (New) The method of claim 108, wherein the plurality of response profiles comprises more than 50 response profiles.

110. (New) The method of claim 109, wherein the plurality of response profiles comprises more than 100 response profiles.

111. (New) The method of claim 38, wherein said first plurality of perturbations is associated with a particular biological effect.

112. (New) The method of claim 111, wherein the particular biological effect is the effect of a particular class or type of drug.

113. (New) The method of claim 111, wherein the particular biological effect is a therapeutic effect.

114. (New) The method of claim 111, wherein the particular biological effect is a toxic effect.

115. (New) The method of claim 38, wherein each of the sets of genes consists of genes which are co-regulated.

116. (New) The method of claim 38, wherein each of the sets of genes consists of genes which co-vary in the plurality of response profiles.

117. (New) The method of claim 116, wherein the genes which co-vary are identified by cluster analysis of genes in the plurality of response profiles.

118. (New) The method of claim 117, wherein the cluster analysis is done by means of a clustering algorithm.

119. (New) The method of claim 118, wherein the clustering algorithm is *hclust*.

120. (New) The method of claim 117, wherein said cluster analysis determines a clustering tree, the genes which co-vary comprising branches of said clustering tree.

121. (New) The method of claim 120, wherein the sets of co-varying genes are selected from a branching level of the clustering tree.

122. (New) The method of claim 117, wherein a statistical significance for the sets of co-varying genes is determined by means of an objective statistical test.

123. (New) The method of claim 122, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in cluster analysis of the genes;

(b) generating permuted response of genes by means of Monte Carlo randomization of perturbation index for the response of each gene across all perturbations;

(c) performing cluster analysis on the permuted response of genes;

(d) determining the fractional improvement in the cluster analysis on the permuted response of genes; and

(e) repeating said steps of generating permuted response of genes and performing cluster analysis on the permuted response of genes so that a distribution of fractional improvements is obtained;

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

124. (New) A method for grouping measured response profiles in sets which are associated with similar biological effects comprising grouping response profiles among a plurality of response profiles into sets, each of said sets of response profiles consisting of response profiles in which the responses of one or more sets of cellular constituents in each response profile are similar among response profiles in the set, each response profile in said plurality of response profiles (i) comprising measurements of a plurality of cellular constituents, and (ii) resulting from a different drug perturbation, wherein each of said sets of cellular constituents consists of cellular constituents that co-vary under a plurality of perturbations or that are co-regulated, wherein said plurality of response profiles comprises at least five response profiles.